A Constrained Combinatorial Register in the Voynich Manuscript: Dual-Zone Placement, a Closed Label Set, and Negative Tests of Richer Readings

Working paper, August 2026. Not a decipherment.

Transcription standard: EVA (IVTFF / Zandbergen–Landini / Stolfi–Takahashi).

ABSTRACT

Label tokens in the Voynich Manuscript are known to differ statistically from running text. This study treats that difference as a functional split rather than as two dialects of one language. Short tokens built on the openings *ot-*, *ok-*, and *ol-* are enriched on and immediately beside drawn elements (tubes, basins, junctions, stem or root sections, circular sector edges). A second register, dominated by the *qok-* family together with *chedy* / *shedy*, occupies free paragraph text and is strongest on the balneological (Bio-B) pages. Residual forms (*daiin* / *aiin*, *dar* / *dal*, *qol*, and other shorts that do not track drawings) are suppressed on clean object labels and rise with distance from the image. A four-band proximity gradient organises placement: object-tied (Band 1) density is highest where text touches the drawing (distance Band A) and lowest in free paragraphs (Band D). After a noise-strip that removes residual, distant prose, unanchored medium text, weak mid-distance tokens, and non-recurrent variants, a thin closed set remains. Those forms are reused across similar object classes and across sections; pairing of two families on one locus is rare. They behave as class or slot markers, not unique names and not prose. Tests of fluent translation, state variables, ordered catalogues, circular numerical templates, rotatable overlays, hidden page-selectors, and independent numerical keys returned null or only weak, non-exclusive leans. The most economical account consistent with the measurements is: drawings first; a small pre-mapped mark set applied by visual proximity; then a large body of text that gives the page the appearance of continuous writing. Dual-zone structure plus the proximity rule are treated as high-confidence features of the manuscript. Expansion of that structure into a publicly accepted plaintext from the pages alone is treated as low-confidence. The project is closed at this granularity unless a new page-level signal or an external historical constraint appears.

1. Introduction

Research on the Voynich Manuscript (Beinecke MS 408) has long distinguished “Labelese” from running text: labels show a flatter frequency distribution, a higher hapax rate, a preference for openings in *o-* (especially *ot-* and *ok-*), and almost no *qo-* openers. That observation is accepted here as background, not as the result.

The working question is narrower. Do the short object-tied tokens and the paragraph text function as two registers of one system, and if so, what placement rule and what surviving token set remain after residual material is removed? The study does not assign English glosses, does not treat EVA strings as phonetic values, and does not promote a form unless it improves the match between text and visual function (drawings or zones). Speculative readings were confined to a sandbox and were not merged into the core.

Two broad interpretations of any remaining structure remain open: an external key that did not travel with the book, or a methodical start followed by a large under-determined remainder. Neither is proven. Both are compatible with the counts.

2. Materials and Method

- **Units.** Locus types follow IVTFF practice: paragraph text (P) versus labels and dislocated words (L), with circular (C) and radial (R) text noted where relevant.
- **Promotion rule.** A form enters the working set only if it improves correspondence with drawings or zones. Residual and medium-register tokens are not promoted on frequency alone.
- **Non-intent filter.** Distributional regularities are recorded as structure. They are not converted into sentences, procedures, or historical motives.
- **Noise-strip.** Five passes removed: (i) residual families; (ii) free-paragraph (Band D) prose; (iii) unanchored medium-register text; (④) weak mid-distance (Bands B and C) tokens; (v) non-recurrent variants. What survives is treated as the closed codebook.
- **Sections sampled.** Hardware-rich Bio-B pages (quire 13), pharmaceutical object pages (cylinders, stems, plant parts), circular diagrams (quires 9–10), zodiac nymph-star labels, mixed herbal pages, and the starred paragraph section as a low-image control.
- **Visual working choice.** Pharmaceutical cylinders were read as stem, stalk, or root sections rather than as apothecary jars. That is a working visual classification, not a botanical identification.

3. Results

3.1 Two registers

- **Band 1 (object / locus register).** Short tokens with openings *ot-*, *ok-*, or *ol-*. Enriched on and next to drawn elements. In this paper "Band 1" names the object-tied token register (*ot-* / *ok-* / *ol-*). Distance from the drawing is written as Bands A–D.
- **Medium register.** *qok-* plus *chedy* / *shedy* and kin. Dominates free paragraph text, especially Bio-B.
- **Residual sieve.** *daiin* / *aiin*, *dar* / *dal*, *qol*, and other shorts that do not behave like Band 1. Suppressed on clean object labels; elevated in distant text.

The split is stronger than a mere dialect contrast. It is treated as two functions plus an unclaimed remainder.

3.2 Proximity gradient

Distance from the drawing is the primary placement rule.

Band	Position	Band 1	Residual / medium
A	Touching the drawing	Highest	Near floor
B	Immediately adjacent	High	Low
C	Moderate gap	Transitional	Rising
D	Free paragraph	Lowest	Highest

The same gradient is obtained measuring outward from the object or inward from the text. Left versus right of a conduit is uninformative. Token length does not track distance. Inlet, mid-pipe, and junction give only a weak secondary lean (extended forms slightly more common at junctions).

3.3 Section ranking

Dual-zone contrast ranks: Bio-B > pharmaceutical pages > circular / later Herbal B > early Herbal A > starred "recipes" pages. The recipes section has few true object-tied Band 1 labels. Marginal stars function as paragraph dividers, not as labelled objects. Residual prominence in running text is high.

3.4 Closed codebook after noise-strip

Surviving families:

- *otol* / *otal* / *otolor*
- *otoldy* / *otaldy*
- *okar* / *okary*
- *okal* / *okalo* (when on hardware)
- *otok* / *otoky*
- *okol* / *okeol*
- short *ol-* cores, especially on Bio-B edges

Observed behaviour:

- One mark per locus. Pairing of two families on one object is rare or absent.
- The same form is reused on multiple objects of similar class or role on one page. These are not unique names.
- Reuse across sections is the rule. Mix shifts by page weight: extended forms are Pharma-heavy; short *ol-* / *otol-* cores are Bio-B-heavy.
- Rare nearby "echoes" occupy Bands B and C and follow the same proximity rule. They are extra instances, not a second function.
- Circular and zodiac labels share the family (*otaldy*, *otoky*, *otaly*, *okary*, *otol*, and kin). That is family overlap, not a proven overlay key.

Working reading of the residue: class, slot, or locus markers placed by proximity on drawn elements, drawn from a closed combinatorial set. Not unique identifiers, states, procedures, or prose.

3.5 Informal lexicon

An informal lexicon (v0.6) was frozen after lock. It is a family list for Band 1, not a translation dictionary. No promotion after lock.

4. Negative Results

Each hypothesis below received a concrete test and returned null or only a weak, non-exclusive lean.

Hypothesis	Result
Fluent translation or fifteenth-century process manual	Interpretive overlay; fails one-to-one EVA lookup
State variables (open/closed, figure-adjacent, flow direction, multiplicity)	Null or weak non-exclusive leans
Label-internal structure as state	Null
Ordered list or catalogue	Weak clustering only; no consistent order

Circular pages as numerical templates matching object counts	No match
Literal geometric overlay or rotatable rings	Hit rates consistent with chance
Circular pages as hidden archive selector or combinatorial key	No coherent sector-label group reappears on a target page
Values on the 1–4 gradient as an independent numerical key	Digitises proximity only; no along-object arithmetic
Inner splits (*ot* / *ok* / *ol*, endings) as a 4 / 8 / 12 / 16 scale	Combinatorial variety only
Inter-choice distances (along object, facing pages, through-object or radial)	Irregular; no third axis
Page-to-page sequence predicting the next page beyond section signature	Null

Remainder after those failures: constrained combinatorial logic, a proximity rule, and local choice from a permitted set. Choice did not yield a second key. Under the keep-rules, most running text classifies as filler.

5. Discussion

5.1 Production rule (minimal account)

A cheaper generative story already explains the counts. No hidden language is required.

- **Pictures first.** Period-legible but largely unidentifiable objects establish loci.
- **Pre-mapped marks.** A small family of short groups is applied to those loci by visual distance. The drawing matters as a grouping base.
- **Fill.** Glyphs—some borrowed from Latin-like bookhands, some invented, including gallows—are used so that the page looks written. That fill need not encode.

The account predicts a tight label layer and a body that will not decode. That is what was observed.

“Constructed curiosity” here means a book assembled to look like a written technical object. It does not require a modern workshop fraud. Carbon-dated parchment and period manufacture are not in dispute. The claim concerns the function of the writing, not a later forgery of the support.

5.2 What the account does not say

- It does not say that every stroke is dummy. Band 1 is too regular.
- It does not assign meanings to *otol* or any other token.
- It does not reopen failed theories. New claims must measure against the baseline.

5.3 Confidence

Dual-zone structure plus the distance gradient are treated as a real feature of the manuscript (approximately 80–85 percent confidence as an internal structural claim). Expansion of that structure into a publicly accepted solution from the pages alone is treated as low (approximately 15–25 percent). The gap is expected if the body was never a language.

6. How the Work May Be Repeated

- Use standard EVA only.
- Promote a form only if it improves match to drawings or zones.
- Do not smooth flags into English sentences.
- Record null results; they are part of the outcome.
- Stop when further pressure only tightens the same floor.

Suggested first check for an independent worker: take one previously unscored Bio-B page and one previously unscored pharmaceutical page; mark only Band A and A/B labels; test whether they fall inside the codebook families. If they do, the baseline still holds. If a new high-purity family appears on hardware, promotion may be considered.

Optional experiment only: a superimposed-pages test. Over-fit risk is high; family overlap is already known.

New work is justified only if it is microscopic and countable against the codebook and distance bands, or if an external constraint appears (quire reconstruction, or an independent historical parallel that maps onto dual-zone structure and the codebook, not onto “plants mean plants”).

7. Conclusion

The measurements that persist are few: two registers, a distance rule, a small reused mark-set on drawings, and a body of text that behaves as fill.

The interpretation that adds the least is: pictures, then group marks, then writing that looks like writing. Coherence of a fuller system was never achieved, or was never required. There is no translation in the sense tested here—only that fixed structure.

Further pressure on the same baseline will not raise the ceiling. The working model is locked at this granularity unless a new page-level signal or an external historical key appears.

One-line remainder: A constrained combinatorial register system whose placement is governed by visual distance, with a thin closed set of class or slot locus marks on drawings, and a large filler remainder in running text. That is the floor. It is not a solve.